

语言和文法

2020年5月6日 18:32

1. 字母表 (有穷符号集合)

- a. 常定义字母表上的连接运算为乘运算, n 次幂运算为 n 个符号连接
- b. 非空串的集合常称为positive closure正闭包, 记作 $+$
- c. 可空串的集合常称为Kleene closure, 记作 $*$
- d. 克林闭包的每个元素都为该字母表的串
- e. 零个终结符组成的空串常记作 ϵ

☐ f. 终结符号串指零个或多个终结符号组成的串

2. 文法/语言规则 $G=(V_T, V_N, P, S)$

- a. terminal symbol终结符是语言的基本符号, 常称作token, 集合为 V_T , 如 a, b, c , 终结符号串如 u, v, w
- b. nonterminal symbol非终结符是表示语法成分的符号, 常称作语法变量, 集合为 V_N , 如 A, B, C
- c. V_T 和 V_N 不相交, 并集为文法符号集, 文法符号如 X, Y, Z , 文法符号串如 α, β
- d. production产生式描述将非终结符和终结符组合成串的方法, 记作 $\alpha \rightarrow \beta$ 读作 α 定义为 β , α 称作head或left side, β 称作body或right side, 集合为 P
 - i. 以某个非终结符为头部的产生式称为该非终结符的产生式
 - ii. 以 ϵ 为体的产生式常称为 ϵ 产生式
- e. 相同左部的产生式可合并成 $\alpha \rightarrow \beta_1 | \beta_2 | \beta_3$ 的形式, 称此时的右部为左部的candidate候选式
- f. S 是start symbol开始符号, 是 V_N 的元素, 默认 P 的首元素的左部为开始符号, 常用 E 表示

3. 语言

- a. derivation推导是通过产生式左侧rewrite替换至右侧, 即抽象到具体, 简单到复杂, 规约reduction是其逆过程
- b. 1步推导directly derive即直接推导, 记作 \Rightarrow , n 步推导记作 \Rightarrow^n , 定义0步推导得到本身
- c. sentential form句型是某文法可推出的文法符号串, 可以为空, 不含非终结符的句型称为sentence句子, 句子的集合即为语言, 记作 $L(G)$
- d. 能规约到文法开始符号的句子就是该文法的语言中的句子, 都可从文法推导出

4. 4类文法 (按从超集到子集的顺序)

- a. 0型文法: unrestricted grammer无限制文法/Phrase Structure Grammer短语结构文法, 只要求产生式左端至少包含一个非终结符
- b. 1型文法: Context-Sensitive Grammer上下文有关文法, 还要求产生式左端长度 \leq 右端 (一般不允许替换出空串)
- c. 2型文法: Context-Free Grammer上下文无关文法, 要求产生式左部必须是非终结符

- d. 3型文法: Regular Grammar正则文法, 分为左/右线性文法, Right Linner右线性文法要求产生式右侧的最右端只能有1个或0个非终结符 (右侧的其余部分是终结符串)
- 5. CFG的分析树
 - a. 根结点是文法开始符号
 - b. 内部结点是产生式
 - c. 叶结点是非终结符或终结符
 - d. 叶结点从左到右排列得到树的yield产出/frontier边缘
- 6. 句型 (符号串) 的短语
 - a. 句型的子树的边缘都是句型的phrase短语
 - b. 只有两代结点的子树的边缘为该句型的immediate phrase直接短语
 - c. 句型直接短语一定是产生式的右部, 反之不成立
- 7. 二义性文法
 - a. 能生成多棵分析树的文法是有ambiguous二义性的
 - b. 判定二义性困难, 但可以给出几个充分条件来保证无二义性
- 8. 一个语法的文法是不唯一的